

Article

Predicting Persistent Forest Fire Refugia Using Machine Learning Models with Topographic, Microclimate, and Surface Wind Variables

Sven Christ ¹ , Tineke Kraaij ² , Coert J. Geldenhuys ³  and Helen M. de Klerk ^{1,4,*} 

¹ Department of Geography and Environmental Studies, Stellenbosch University, P. Bag X1, Stellenbosch 7602, South Africa

² Natural Resource Science and Management, Science Faculty, Nelson Mandela University, George Campus, George 6530, South Africa; tineke.kraaij@mandela.ac.za

³ Forest Science Program, Department of Plant and Soil Sciences, University of Pretoria, c/o Forestwood cc, 35 Grace Avenue, Murrayfield, Pretoria 0184, South Africa; cgelden@mweb.co.za

⁴ Centre for Geospatial and Computing Technologies, Faculty of Environment, Society and Design | Te Wāhaka ki te Taiao, te Hāpori Whānui me kā Mahi Hoahoa, Lincoln University, Forbes Building, Ellesmere Junction Road, Lincoln 7608, New Zealand

* Correspondence: helen.deklerk@lincoln.ac.nz

Abstract

Persistent forest fire refugia are areas within fire-prone landscapes that remain fire-free over long periods of time and are crucial for ecosystem resilience. Modelling to develop maps of these refugia is key to informing fire and land use management. We predict persistent forest fire refugia using variables linked to the fire triangle (aspect, slope, elevation, topographic wetness, convergence and roughness, solar irradiation, temperature, surface wind direction, and speed) in machine learning algorithms (Random Forest, XGBoost; two ensemble models) and K-Nearest Neighbour. All models were run with and without ADASYN oversampling and grid search hyperparameterisation. Six iterations were run per algorithm to assess the impact of omitting variables. Aspect is twice as influential as any other variable across all models. Solar radiation and surface wind direction are also highlighted, although the order of importance differs between algorithms. The predominant importance of aspect relates to solar radiation received by sun-facing slopes and resultant heat and moisture balances and, in this study area, the predominant fire wind direction. Ensemble models consistently produced the most accurate results. The findings highlight the importance of topographic and microclimatic variables in persistent forest fire refugia prediction, with ensemble machine learning providing reliable forecasting frameworks.

Keywords: fire refugia; persistent forest fire refugia; machine learning; ensembles; random forest; extreme gradient boosting (XGBoost); aspect; topography; microclimate; surface wind



Academic Editors: Wolfgang Kainz and Godwin Yeboah

Received: 7 September 2025

Revised: 24 November 2025

Accepted: 29 November 2025

Published: 5 December 2025

Citation: Christ, S.; Kraaij, T.; Geldenhuys, C.J.; de Klerk, H.M. Predicting Persistent Forest Fire Refugia Using Machine Learning Models with Topographic, Microclimate, and Surface Wind Variables. *ISPRS Int. J. Geo-Inf.* **2025**, *14*, 480. <https://doi.org/10.3390/ijgi14120480>

Copyright: © 2025 by the authors. Published by MDPI on behalf of the International Society for Photogrammetry and Remote Sensing. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Forest fire refugia are fire-free areas within fire-prone landscapes (persistent fire refugia, [1,2]). Sometimes fire refugia are defined as areas that experience fire less frequently or with lower severity than their surroundings (ephemeral fire refugia, [1,3,4]). Forest fire refugia are often surrounded by forest, but sometimes they are imbedded in other vegetation types, such as Mediterranean shrublands [5,6] or Spanish open woodlands [2]. Fire refugia are ecologically important, providing habitat for fire-sensitive forest flora and

fauna in fire-prone matrices, shelter during the fire, and food after the fire, as well as maintaining mature vegetation and seed sources, which are essential for post-fire forest ecosystem recovery. Consequently, forest fire refugia enhance forest resilience [7].

Fire behaviour and spread is determined by factors illustrated by the fire behaviour triangle, namely, fuels, weather, and topography [1,8]. For a fire to start there must be fuel to burn, oxygen, and an ignition, often termed 'heat'. Once started, fire behaviour will be determined by types and condition of the fuel, weather, and topographical features. Dry fuel promotes fire spread, whereas wet, discontinuous, or sparse fuel slows or stops fire. Hot dry windy weather will promote fire spread, whereas cold, damp weather can help slow or contain a fire. Topographical features such as slope or channels can promote fire spread. Cooler slopes and drainage lines can slow a fire. Scree slopes and cliffs can be a barrier to fires [9]. Topographic features, such as hills, mountains, and deep valleys, can shield refugia from fire driven by prevailing winds, particularly hot and dry foehn and katabatic winds [2]. All of these factors have been shown to be important in the formation of fire refugia in North American forests [3,10,11] and Mediterranean-type ecosystems [2]. Understanding the interplay between topography, surface wind movement, local climatic conditions, and fire behaviour is essential for identifying and conserving persistent forest fire refugia. No management technique can completely ensure that refugia will not burn under extreme fire conditions. However, knowledge of forest fire refugia locations and drivers can inform forest management practices aimed at promoting ecosystem resilience in fire-prone regions (*sensu* [9]), such as ensuring that these areas are not harvested.

Studies aiming to identify predictors of forest fire refugia have examined slope, aspect, the topographic wetness index, the topographic convergence index, topographic roughness or ruggedness [3,9,12], and elevation [13]. The topographic wetness index is a measure of water accumulation and soil saturation and based on landscape position [14]. The topographic convergence index is a metric of cold air pooling [3]. Topographic roughness is a measure of topographic complexity, and solar irradiation relates to heat received by a site.

Globally, forest occurrence is determined by climate and soil characteristics and fire exposure, with feedback loops of shading and moisture availability due to forest presence further promoting the location of forest patches [5,15–19]. Afromontane forests in the Southern Cape and Tsitsikamma region of South Africa occur in a matrix of fire-prone fynbos shrublands in locations protected from fire [5,20,21] and with higher soil moisture [16,17,22]. Typically, fire-protected positions in this landscape occur on the lee side of topographic features in relation to the directions of desiccating katabatic ('berg') winds [5,19,23,24]. In the Southern Cape region, forest refugia seldom burn, with patches recorded as surviving fire in excess of 100–230 years [25,26], while some patches never burn [27]. The fire regime of surrounding fynbos sees a mean fire return interval of 10–13 years [28,29].

This study will include surface wind speed and direction derived from computational fluid dynamics models, along with various topographic and microclimatic variables, in machine learning models to predict persistent forest fire refugia locations in the Southern Cape of South Africa to inform management action planning.

2. Materials and Methods

2.1. Study Area

The Southern Cape and Tsitsikamma region of South Africa (running from Stormsriver in the east to George in the west, Figure 1) contains the largest area of Afrotropical forest in southern Africa [19,30]. The forest occurs embedded in fire-prone, fire-adapted fynbos, which is one of the five Mediterranean-type ecosystems of the world and part of the Cape Floristic Region [31]. Boundaries between forest patches

and surrounding fynbos vegetation are likely determined by fire occurrence, climate, and soil characteristics [5,15–19]. On a local scale, the potential of a landscape to accommodate forest is shaped by site characteristics, including climatic and substrate variables. In the relatively uniform Tsitsikamma region, evergreen forest is supported across a rainfall gradient of over 1200 mm/year to about 500 mm/year, a pattern consistent across South Africa and southeastern Democratic Republic of Congo [32]. Detailed field work by one of the authors (250 plots within evergreen forests along six transects from mountain to coast in the Southern Cape) recorded species composition and substrate variables [33]. In all areas, fire-adapted fynbos shrublands occurred adjacent to evergreen forest along typical fire pathways, with slope as the key differentiating factor, supporting the hypothesis that fire dynamics convert landscape potential into actual vegetation patterns, with forests persisting in fire-shadow zones and fire-adapted communities dominating fire pathways.

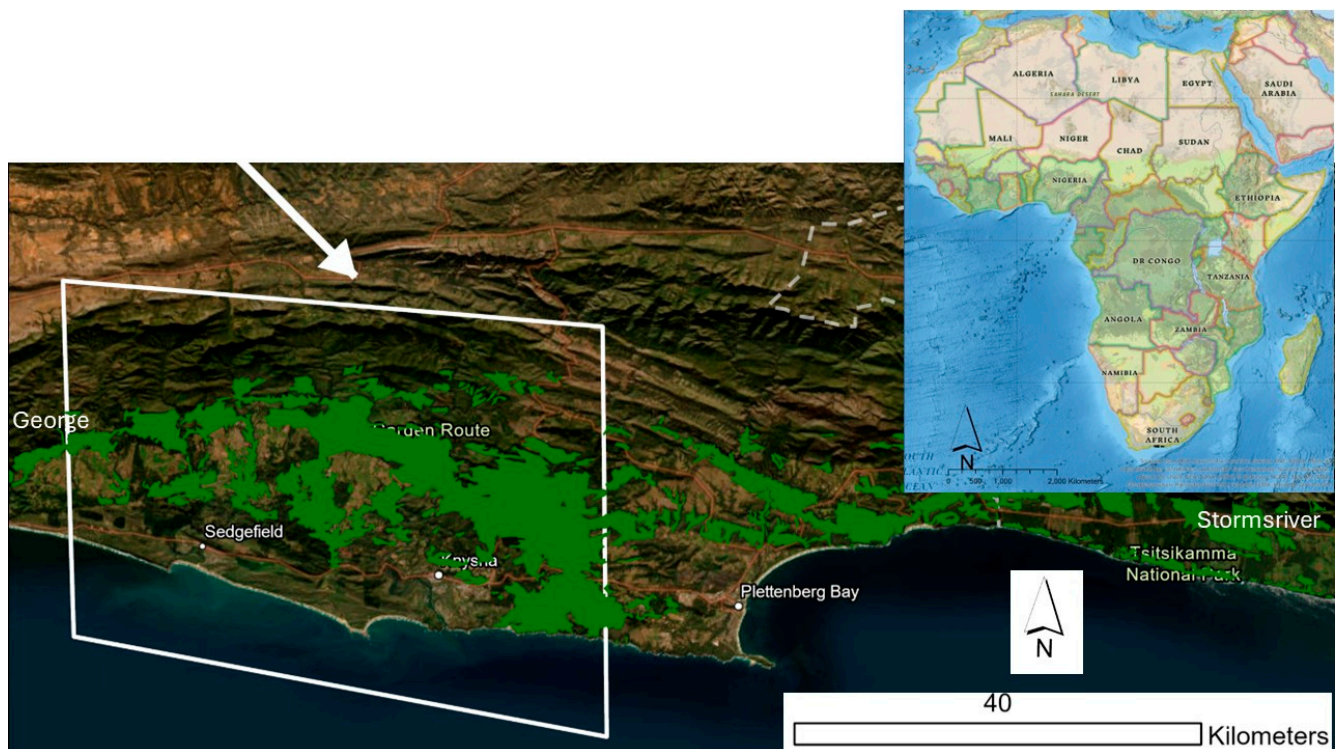


Figure 1. The Sedgefield–Knysna area (white box) within the Southern Cape–Tsitsikamma region of South Africa experiences northwesterly foehn winds (indicated by the white arrow) as the predominant fire weather wind. Afrotropical forest of the region is shown in dark green.

The narrow (10–40 km wide) coastal plain of the Southern Cape–Tsitsikamma region has a predominantly southerly aspect and rises from the Indian Ocean in the south to the foothills and rugged crest of the Outeniqua and Tsitsikamma Mountains in the north, peaking at 1675 m [34]. Due to the maritime influence, the climate is temperate. Although rain falls throughout the year, it peaks in spring and autumn. Summer months are the driest due to high evaporative demand, particularly associated with prevailing southeasterly winds. In winter, hot and dry northerly bergwinds sporadically create high-fire-danger conditions conducive to fires [34,35]. Consequently, fires may occur at any time of the year [29], and there is no distinct fire season in the study area. The western Sedgefield–Knysna portion of the Southern Cape is the focus of this study. It experiences northwesterly foehn winds (white box, Figure 1).

2.2. Data for Persistent Forest Fire Refugia

The fire regime of the forest is characterised by very long fire return intervals and patches that never burn [27], with an analysed tree ring showing that at least one forest patch has remained unburned for over 230 years [25]. Persistent forest fire refugia were thus identified as areas not burned in the last 100 years [34] and which have forest stands in contrast to the surrounding fynbos matrix, which is a shrubland of smaller stature and which burns at fire return intervals of 10–13 years [27,36–38]. This definition is similar to the analysis in [26]. The 100-year threshold used to define persistent refugia is ecologically grounded but remains somewhat arbitrary and could be refined through sensitivity testing. In addition, the models are based on current climatic and fire regime conditions. Persistent forest fire refugia were initially identified using the MODIS burned area product (500 m) between 2000 to 2022 [36]. These were subsequently validated through manual interpretation of aerial photographic imagery with a pixel size of 10–20 m and spanning approximately 100 years (1930–2022). This provided finer spatial detail and extended temporal coverage, helping to mitigate the under-detection of small or low-intensity fires that may be missed by coarse satellite products of only two decades. This combined approach improves confidence in refugia mapping, though it does not fully eliminate uncertainty associated with temporal and spatial data gaps. Presence of persistent forest stands is visible on all imagery and was verified by local experts with thorough in-field knowledge of the forest stands. This set of persistent forest fire refugia was iteratively subjected to a 75/25 split for training and validation.

2.3. Data for Predictor Variables

Human-impacted areas such as urban developments and intensively farmed areas were excluded using the 2020 South African National Landcover Classification dataset (<https://za.africageoportal.com/maps/cff4e27fa6fb46f6bde59edb889398ef/explore>; accessed 15 June 2023). We investigated the following topographic, microclimatic, and surface wind direction and speed variables that can be easily measured over large areas in a consistent manner as possible correlates of the location of persistent forest fire refugia (Table 1). The Shuttle Radar Topographic Mission's (SRTM; <https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1>, USGS, 2004; accessed 4 February 2023) Digital Elevation Model (DEM) [37,38] is supplied at a spatial resolution of 30 m. The SRTM specifications indicate an absolute vertical accuracy of ± 5.94 to ± 16 m at the 95% confidence level [39]. The SRTM DEM has been thoroughly tested [40,41] and was used to create variables for elevation [13], slope, aspect, the topographic wetness index, the topographic convergence index [3,42], and the topographic roughness index or ruggedness [5,12]. Processing was conducted in QGIS (3.18.2).

Long-term averages of local-scale annual temperature and annual solar irradiation (kWh/m^2) were sourced from the World Bank's Global Solar Atlas at a 250 m spatial resolution [43] (<https://solargis.com>). These data represent values at 2 m above ground level [43]. These were used in preference to the calculation of solar irradiation from a DEM, which would still need to be validated and is, at best, an estimation of the potential solar irradiation [44]. Global radiation, sometimes called global horizontal irradiation, is the total amount of solar radiation reaching a horizontal surface. It consists of direct radiation (or direct normal irradiation) that reaches the surface and diffuse radiation (diffuse horizontal irradiation) that is scattered by the atmosphere. We include all three in the models.

Surface wind direction and speed were calculated using Computational Fluid Dynamics (CFD) in OpenFOAM [45] for 30 m pixels. Wind speed and direction were calculated at a height of 2 m above the surface—an elevation critical for fire modelling and behaviour [46]. Internally the simulation achieved stable convergence, with the k-equation

residual reducing from 5.54×10^{-4} to 3.41×10^{-5} , and global and cumulative continuity residuals were -8.53×10^{-10} and -1.01×10^{-4} , respectively, confirming good mass conservation. As no meteorological data points were available within the study area, a pattern-based validation approach was used to evaluate CFD wind direction and speed outputs against known wind movements *sensu* [47] (Appendix S1) and previously recorded fire paths (Appendix S2). The latter showed that the fire spread model that included the CFD-derived wind data layers results most accurately matched the recorded [48] 45 min from the fire ignition point to the arrival of the fire at the Kooigoed weather station (Appendix S2). This type of validation, whereby patterns produced by model output results are compared with patterns observed in the real world, is often used for evaluating the validity of models in agent-based modelling in ecology, another type of modelling in which it is frequently difficult to have exact validation data [49]. Although this provides confidence in directional accuracy and fire spread impact, future work would benefit from additional field-based wind measurements to refine model calibration. Wind speed and direction were converted into raster datasets, georeferenced, and interpolated using neighbourhood averaging to fill gaps in QGIS.

All datasets were resampled using bilinear interpolation to a common, co-registered, aligned 30 m grid based on the SRTM DEM in Python 3.10.9. Bilinear interpolation has been shown to be an effective method for continuous datasets and DEM data [50]. While the original spatial resolution of the Solargis datasets should be considered in interpretations, upsampling allowed for retention of crucial topographic details that would otherwise be lost [51].

Table 1. Variables used to model spatial locations of persistent forest fire refugia locations.

Variable	Relevance	Reference
Elevation	Higher elevations may be cooler: >moisture <flammable Some sites also see the following: <fuel continuity at highest elevations	[3,4,13,14]
Slope	Steeper slopes: >fuel preheating >updrafts >fire spread	[3,9]
Aspect	Sun-facing: >solar radiation <moisture Opposite aspect to prevailing fire season wind direction: fire shield	[3,9]
Topographic wetness index	Water accumulation: >moisture <flammability	[3]
Topographic convergence index	Cold air pooling: >moisture <flammability	[3]
Topographic roughness index	Rugged terrain: >volatile surface wind > complex fire behaviour >fire spread	[5]

Table 1. Cont.

Variable	Relevance	Reference
Temperature	Hot areas: <moisture >flammability	
Solar irradiation (global, direct, and diffuse)	>heat <moisture >flammability	[4]
Wind direction	Prevailing fire wind direction: >heat <moisture >fire spread Lee of prevailing wind direction may experience fire skipping	[46]
Wind speed	>wind speed <moisture >flammability and >pre-heating >rate of fire spread	[46]

2.4. Machine Learning Models

Several machine learning techniques, chosen for their differing assumptions and suitability for spatial datasets, were employed in this study. To enhance model performance and reduce overfitting, the dataset underwent Principal Component Analysis (PCA), which improved separability between forest fire refugia and non-refugia areas [52]. PCA1 and PCA2 explained 68.03% and 20.38% of variance, respectively, and were used to increase separability of refugia and non-refugia (Appendix S3). This refinement resulted in 9959 forest refugia pixels and 21,769 non-refugia pixels for training. As there are more than double the number of non-refugia pixels compared to forest refugia pixels in the training dataset, the training dataset can be considered imbalanced. Imbalanced datasets often cause inaccurate results in classifications [53], and consequently, adaptive synthetic (ADASYN) oversampling was tested as it has been found to be useful in improving classification with imbalanced training data [54]. ADASYN aims to keep classes truer to the original data while addressing class imbalance issues. To further reduce overfitting, a 70/30 spatial block holdout split was used for training and validation. This was performed by splitting the input raster into 64x64 pixel blocks and assigning each an ID-only block that contains refugia or non-refugia that are considered for the training or validation. All machine learning algorithms were executed 10 times with different random seeds for the block selection, and the reported results represent the mean, standard deviation, and 95% confidence intervals across these spatial differences. These splits are run iteratively as specified after the discussion on GridSearch.

Three machine learning algorithms—Random Forest (RF), Extreme Gradient Boosting (XGBoost), and K-Nearest Neighbours (KNN)—were applied to assess the relationship between variables described earlier and forest fire refugia and to predict unknown forest fire refugia. RF is an ensemble learning method that constructs multiple decision trees and aggregates their outputs [55]. This approach reduces overfitting and enhances generalisability [56]. RF also provides feature importance rankings, aiding the understanding of key drivers of forest fire refugia [55]. The classifier, implemented in Python using Scikit-learn (Appendix S4), is effective at handling correlated datasets [57]. XGBoost is

a decision-tree-based algorithm that iteratively improves predictions through gradient boosting [58]. It is well-regarded for its speed, scalability, and effectiveness in handling imbalanced datasets [59]. XGBoost was implemented using the XGBoost package in Python (Appendix S5), with precautions taken to mitigate overfitting [60]. K-Nearest Neighbours (KNN) is a non-parametric algorithm that classifies data points based on proximity. It does not require large training datasets and can perform well with both balanced and imbalanced datasets [61]. However, it is sensitive to the choice of K and may experience performance degradation with correlated features [62]. The KNN implementation was performed in Python (Appendix S6). To enhance classification performance, ensemble learning was applied using both hard voting (majority vote, Appendix S7) and soft voting (weighted vote, Appendix S8) approaches. Ensemble methods leverage the strengths of multiple classifiers to improve predictive accuracy. Similar sets of algorithms have been used in other recent studies looking at factors affecting various properties of fire [63,64].

Using default parameters for machine learning often gives suboptimal results [65]. Fine-tuning of parameters, such as the minimum number of samples per split, the minimum samples per leaf, number of estimators, maximum number of features, and maximum depth, can improve model results by up to 40% [65,66]. Grid search, a widely used hyperparameter optimisation technique for machine learning algorithms, searches through a predefined set of hyperparameter values and evaluates the metrics such as r2 and accuracy normally using k-fold. A variety of settings were tested for each technique using grid search (Table 2). The selected setting identified by the grid search as optimal for each experiment can be seen in Appendix S9. Grid search was run for the first experiment for each of RF, XGBoost, and KNN. The remaining nine experiments applied these settings using a stable random seed for the classifier and a varying random seed for the training/validation dataset. The ensemble methods used the best parameters from the single-method experiments with the same features. The number of iterations for cross-validation (random splitting of the training and verification datasets) run for each of the 12 experiments were calculated as the product of the number of possible combinations of the GridSearchCV parameters and the GridSearch folds (CV) which was set to 3 and totalled 1296 for RF, 2160 for XGBoost, and 72 for KNN. GridSearchCV in Scikit-learn performs k-fold cross-validation.

A total of 30 machine learning experiments were conducted, systematically varying dataset inputs and employing ADASYN oversampling in subsequent runs. Within each of the five machine learning experiments, variables were omitted to determine the impact of their exclusion, starting with exclusion of elevations (experiments 3, 9, 15, 21, and 27, Appendix S10), temperature (experiments 4, 10, 16, 22, and 28), diffuse irradiation, direct irradiation, and global irradiation (experiments 5, 11, 17, 23, and 29), and wind speed and direction, until only DEM-derived variables (aspect, slope, topographical convergence, topographical roughness, and topographical wetness index) remain (experiments 6, 12, 18, 24, and 30).

Table 2. Gridsearch hyperparameterisation settings tested and optimisation settings selected (bold).

Parameter	Search Values			
Random Forest				
Out-of-bag score	TRUE			
N estimators	50	100	200	500
Max features	sqrt	log2		
Max depth	none	10	20	30
Min samples split	2	5	10	
Min samples leaf	1	2	4	

Table 2. Cont.

Parameter	Search Values				
XGBoost					
Parameter search values					
Eta (Learning rate)	0.01	0.1	0.5		
N estimators	50	100	200	1500	
Max depth	3	5	7		
Reg_alpha (Lasso regression)	0	0.001	0.005	0.01	0.05
Reg_lambda (Ridge regression)	0.5	1	1.5	2	
KNN					
Parameter search values					
N neighbours	3	5	7	9	
Weights	uniform	distance			
Metric	Euclidean	Manhattan	Minkowski		

2.5. Model Performance and Evaluation

Standard accuracy metrics were used, including accuracy, precision, and recall. These measures of fit are frequently used in the evaluation of machine learning models and provide a multi-faceted approach to evaluating the model fit. Feature importance was used for the RF models, and gain, which can be interpreted as a form of feature importance, was used in XGBoost. Standard deviation and 95% confidence intervals were calculated for feature importance and gain over the multiple runs.

3. Results

3.1. What Determines Persistent Forest Fire Refugia

Aspect is consistently the most important variable identified for modelling persistent forest fire refugia locations based on both feature importance and gain (Figure 2, Table 3). It is identified as twice as important as any other variable. Solar irradiation, particularly global horizontal irradiation, is also identified as important across all models (Figure 2, Table 3). Surface wind direction and speed have the next largest impact on feature importance and gain (Table 3), although the exact order of importance differs between algorithms. Temperature is only identified as important by Random Forest models. Temperature showed some importance in XGBoost (experiment 7, Table 3), but this was reduced once ADASYN oversampling was implemented (experiment 8, Table 3). For the other classifiers, removing temperature had a small impact (experiments 4, 10, 16, 22, and 28), with a few more forest fire refugia being predicted, some of which were shown to house forests when inspected on high-resolution imagery. Overall, RF models spread feature importance more broadly, with XGBoost concentrating mostly around aspect. While error bars show more variability over the experiment runs around aspect than the other features, aspect still remains the primary feature of importance for both RF and XGBoost.

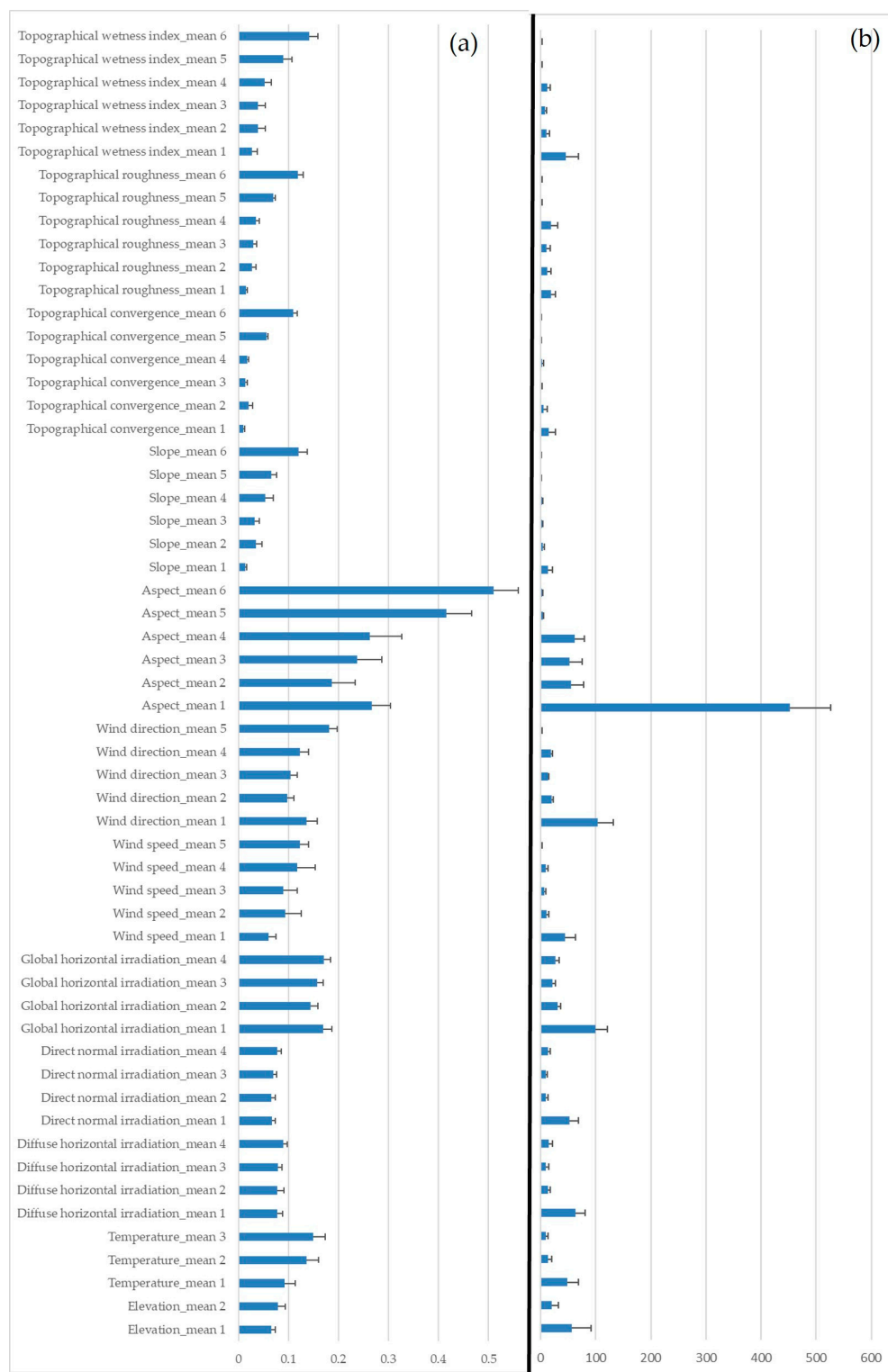


Figure 2. Feature importance (a) from Random Forest experiments (values range from 0 to 1), and feature gain (b) from XGBoost (XGBoost, values are unbounded) showing the relative contribution of each variable to model performance.

Table 3. Model feature importance (RF, values range from 0 to 1) and gain (XGBoost, values are unbounded) results. Strongest values are in the darkest shades of green.

Method	Experiment Number	Elevation	Temperature	Diffuse horizontal Irradiation	Direct Normal Irradiation	Global Horizontal Irradiation	Wind Speed	Wind Direction	Aspect	Slope	Topographical Convergence	Topographical Roughness	Topographical Wetness Index
RF	1	0.07	0.09	0.08	0.07	0.17	0.06	0.14	0.27	0.01	0.01	0.02	0.03
	2	0.08	0.14	0.08	0.07	0.14	0.09	0.10	0.19	0.04	0.02	0.03	0.04
	3		0.15	0.08	0.07	0.16	0.09	0.10	0.24	0.03	0.01	0.03	0.04
	4			0.09	0.08	0.17	0.12	0.12	0.26	0.05	0.02	0.04	0.05
	5						0.12	0.18	0.42	0.07	0.06	0.07	0.09
	6								0.51	0.12	0.11	0.12	0.14
XGBoost	7	56.27	48.56	63.49	52.52	99.87	44.88	103.19	452.29	13.23	14.88	18.83	45.64
	8	19.66	13.67	13.07	10.02	30.49	10.33	19.57	55.02	3.99	5.32	12.69	10.44
	9		9.61	9.94	9.02	21.65	7.07	13.28	52.39	2.29	1.52	11.27	7.92
	10			14.96	13.08	27.28	9.54	18.84	62.25	3.21	3.19	19.18	12.26
	11						0.83	1.72	4.62	0.30	0.58	1.06	0.92
	12								2.31	0.50	0.56	0.72	0.67

3.2. Model Performance

Accuracy metrics suggest that the results of all experiments (all algorithms/models across all combinations of variables) can be treated with confidence with an average overall accuracy of 0.82 (Table 4). The ensemble methods and XGBoost performed the best, with RF showing slightly lower accuracy. KNN had the lowest performance, especially in recall. The unbalanced datasets (with no ADASYN oversampling) generally performed well across all classifiers but struggled with recall. The best-performing experiments overall, especially in recall, were those where elevation and temperature were dropped with all other variables kept. Comparing the mapped outputs of experiments RF3 and 4 with KNN15 and 16 (Figure 3), it is clear that KNN predicts more lowland forests, some of which may be refugia, but others are overestimates when compared to aerial imagery. KNN, however, dramatically overestimates forest fire refugia once elevation is removed (experiments 3, 9, and 15, Figure 3), indicating that KNN outputs are sensitive to which variables are included in the model.

Table 4. Model performance for various experiments to evaluate the effect on accuracy of omitting certain variables. For each accuracy metric, the mean, standard deviation (Std), and upper and lower 95% confidence intervals (CI_low and CI_high) are presented. Strongest values are in the darkest shades of green. Top scoring model per measure (as measured by the mean) is highlighted in bold.

Method	Experiment_No	Accuracy_Mean	Accuracy_Std	Accuracy_CI_Low	Accuracy_CI_High	Precision_Mean	Precision_Std	Precision_CI_Low	Precision_CI_High	Recall_Mean	Recall_Std	Recall_CI_Low	Recall_CI_High	F1_Mean	F1_Std	F1_CI_Low	F1_CI_High	OOB_Mean	OOB_Std	OOB_CI_Low	OOB_CI_High	AUC_ROC_Mean	AUC_ROC_Std	AUC_ROC_CI_Low	AUC_ROC_CI_High
RF	1	0.810	0.097	0.750	0.870	0.796	0.146	0.706	0.886	0.624	0.257	0.464	0.783	0.654	0.185	0.540	0.769	0.997	0.001	0.996	0.997	0.902	0.068	0.861	0.944
RF	2	0.813	0.088	0.758	0.867	0.794	0.142	0.706	0.882	0.623	0.228	0.481	0.764	0.660	0.166	0.557	0.763	0.998	0.000	0.998	0.999	0.895	0.067	0.853	0.936
RF	3	0.822	0.082	0.771	0.873	0.788	0.141	0.700	0.875	0.664	0.229	0.522	0.806	0.685	0.157	0.587	0.782	0.998	0.000	0.998	0.999	0.908	0.058	0.872	0.944
RF	4	0.834	0.045	0.806	0.862	0.762	0.163	0.661	0.863	0.707	0.136	0.623	0.791	0.716	0.110	0.648	0.784	0.993	0.002	0.992	0.995	0.914	0.028	0.896	0.932
RF	5	0.831	0.030	0.812	0.849	0.716	0.152	0.622	0.810	0.773	0.115	0.702	0.845	0.729	0.096	0.670	0.789	0.965	0.004	0.963	0.968	0.914	0.027	0.898	0.931
RF	6	0.806	0.043	0.779	0.833	0.662	0.157	0.564	0.759	0.836	0.067	0.794	0.878	0.723	0.094	0.665	0.782	0.891	0.014	0.882	0.900	0.894	0.031	0.875	0.913
XGboost	7	0.821	0.089	0.766	0.876	0.766	0.129	0.686	0.846	0.685	0.238	0.537	0.833	0.692	0.156	0.595	0.789					0.905	0.058	0.869	0.940
XGboost	8	0.820	0.084	0.768	0.871	0.761	0.131	0.680	0.842	0.682	0.220	0.546	0.819	0.691	0.145	0.601	0.782					0.897	0.061	0.859	0.935
XGboost	9	0.825	0.082	0.774	0.876	0.755	0.128	0.676	0.834	0.707	0.223	0.569	0.845	0.703	0.145	0.614	0.793					0.898	0.068	0.856	0.940
XGboost	10	0.843	0.039	0.819	0.868	0.758	0.152	0.664	0.852	0.743	0.119	0.670	0.817	0.739	0.101	0.676	0.801					0.920	0.031	0.902	0.939
XGboost	11	0.819	0.023	0.805	0.833	0.709	0.153	0.614	0.804	0.734	0.107	0.668	0.801	0.707	0.092	0.650	0.764					0.895	0.027	0.879	0.912
XGboost	12	0.796	0.036	0.774	0.819	0.662	0.156	0.565	0.758	0.775	0.062	0.737	0.814	0.699	0.087	0.645	0.753					0.867	0.030	0.848	0.885
KNN	13	0.812	0.043	0.785	0.839	0.709	0.163	0.608	0.810	0.741	0.133	0.659	0.823	0.703	0.092	0.646	0.761					0.842	0.045	0.814	0.869
KNN	14	0.804	0.039	0.780	0.827	0.685	0.171	0.579	0.791	0.758	0.123	0.681	0.834	0.698	0.101	0.635	0.761					0.822	0.042	0.796	0.848
KNN	15	0.823	0.026	0.807	0.839	0.693	0.158	0.595	0.791	0.792	0.100	0.730	0.854	0.726	0.106	0.660	0.791					0.844	0.032	0.824	0.863
KNN	16	0.821	0.027	0.804	0.837	0.690	0.158	0.592	0.788	0.792	0.100	0.730	0.854	0.724	0.105	0.659	0.789					0.842	0.031	0.823	0.861
KNN	17	0.799	0.027	0.782	0.815	0.656	0.151	0.562	0.750	0.786	0.087	0.732	0.839	0.701	0.091	0.644	0.757					0.834	0.023	0.820	0.849
KNN	18	0.778	0.036	0.756	0.801	0.630	0.153	0.535	0.725	0.782	0.058	0.746	0.818	0.682	0.091	0.626	0.739					0.828	0.024	0.813	0.843
EnsembleS	19	0.818	0.087	0.763	0.872	0.771	0.137	0.687	0.856	0.671	0.234	0.526	0.816	0.683	0.154	0.588	0.779					0.909	0.056	0.874	0.944
EnsembleS	20	0.823	0.078	0.774	0.871	0.762	0.145	0.672	0.852	0.695	0.212	0.564	0.827	0.698	0.140	0.611	0.785					0.906	0.056	0.871	0.941
EnsembleS	21	0.835	0.070	0.792	0.878	0.764	0.132	0.682	0.845	0.726	0.197	0.603	0.848	0.721	0.127	0.643	0.800					0.919	0.040	0.894	0.944
EnsembleS	22	0.843	0.036	0.821	0.865	0.745	0.163	0.644	0.846	0.765	0.117	0.692	0.837	0.741	0.111	0.673	0.810					0.919	0.026	0.903	0.935
EnsembleS	23	0.825	0.027	0.809	0.842	0.702	0.154	0.606	0.798	0.780	0.108	0.714	0.847	0.725	0.096	0.666	0.785					0.908	0.026	0.892	0.924
EnsembleS	24	0.803	0.043	0.776	0.830	0.658	0.158	0.560	0.756	0.833	0.065	0.793	0.873	0.720	0.097	0.660	0.780					0.890	0.030	0.872	0.908

Table 4. Cont.

Method	Experiment_No	Accuracy_Mean	Accuracy_Std	Accuracy_CI_Low	Accuracy_CI_High	Precision_Mean	Precision_Std	Precision_CI_Low	Precision_CI_High	Recall_Mean	Recall_Std	Recall_CI_Low	Recall_CI_High	F1_Mean	F1_Std	F1_CI_Low	F1_CI_High	OOB_Mean	OOB_Std	OOB_CI_Low	OOB_CI_High	AUC_ROC_Mean	AUC_ROC_Std	AUC_ROC_CI_Low	AUC_ROC_CI_High
EnsembleH25		0.814	0.092	0.757	0.871	0.777	0.140	0.690	0.864	0.653	0.240	0.504	0.802	0.673	0.163	0.572	0.774					0.909	0.056	0.874	0.944
EnsembleH26		0.820	0.082	0.769	0.871	0.771	0.141	0.683	0.859	0.672	0.219	0.536	0.807	0.687	0.147	0.596	0.778					0.906	0.056	0.871	0.941
EnsembleH27		0.831	0.075	0.785	0.877	0.771	0.134	0.688	0.854	0.708	0.212	0.577	0.839	0.711	0.137	0.626	0.796					0.919	0.040	0.894	0.944
EnsembleH28		0.841	0.038	0.818	0.865	0.750	0.165	0.648	0.853	0.750	0.123	0.673	0.826	0.736	0.113	0.665	0.806					0.919	0.026	0.903	0.935
EnsembleH29		0.828	0.028	0.810	0.845	0.708	0.155	0.612	0.804	0.775	0.111	0.706	0.844	0.726	0.099	0.665	0.787					0.908	0.026	0.892	0.924
EnsembleH30		0.805	0.045	0.777	0.833	0.659	0.159	0.560	0.758	0.840	0.063	0.800	0.879	0.723	0.099	0.662	0.785					0.890	0.030	0.872	0.908
Mean		0.819				0.728				0.736				0.706				0.974				0.891			
Std dev		0.014				0.048				0.059				0.022				0.039				0.030			
Min		0.778				0.630				0.623				0.654				0.891				0.822			
Max		0.843				0.796				0.840				0.741				0.998				0.920			

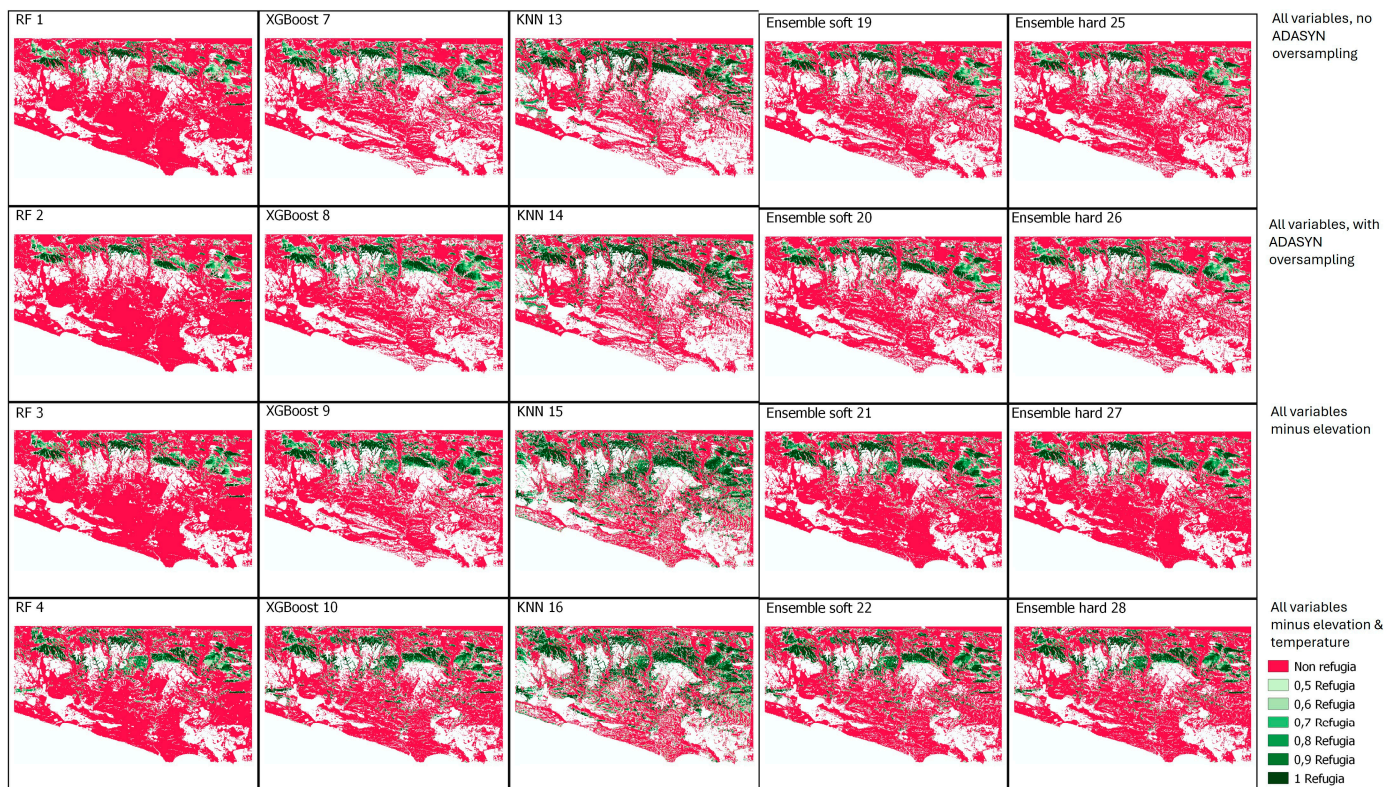


Figure 3. Spatial arrangement of predicted persistent forest fire refugia (green) and non-refugia (red) for classifications using RF, XGBoost, KNN, ensemble with soft voting, and ensemble with hard voting across various experiments.

4. Discussion

4.1. What Determines Forest Fire Refugia

Aspect has been highlighted by other studies as an important variable in predicting persistent forest fire refugia. Previous studies [3,13], found strong links between aspect and fire exclusion, which is what we term persistent forest fire refugia. Aspect of a slope can relate to the solar radiation received and the energy or heat balance and the resultant moisture balance. This is also demonstrated by the decrease in model accuracy when solar irradiation is removed. Sun-facing slopes have higher solar radiation [13], less moisture, and drier, more flammable vegetation or fuels [9] (Figure 4). In the southern hemisphere, these are the north and northwestern aspects, whereas in the northern hemisphere they are the southern slopes. Aspect can also affect the type of vegetation or fuel and how continuous it is, with warmer slopes sometimes having sparser vegetation with less fuel continuity and also different vegetation types that are adapted to hotter, drier conditions, which often result in vegetation structures that are more flammable [67]. In this study area, hot northwestern slopes are also those facing the predominant fire wind direction, and cool southwestern slopes act as fire shields that are very important for forest fire refugia persistence [5]. Persistent long-term forest fire refugia are seen in deep valleys on the south-to-southeast aspect in Figure 5. Similar spatial patterns of evergreen forest persistence within fire-shadow zones have been observed throughout southern and tropical Africa, including the Congo Basin, Mount Cameroon, Ethiopia, Malawi, Mozambique, Angola, and across South Africa's varied landscapes [32].

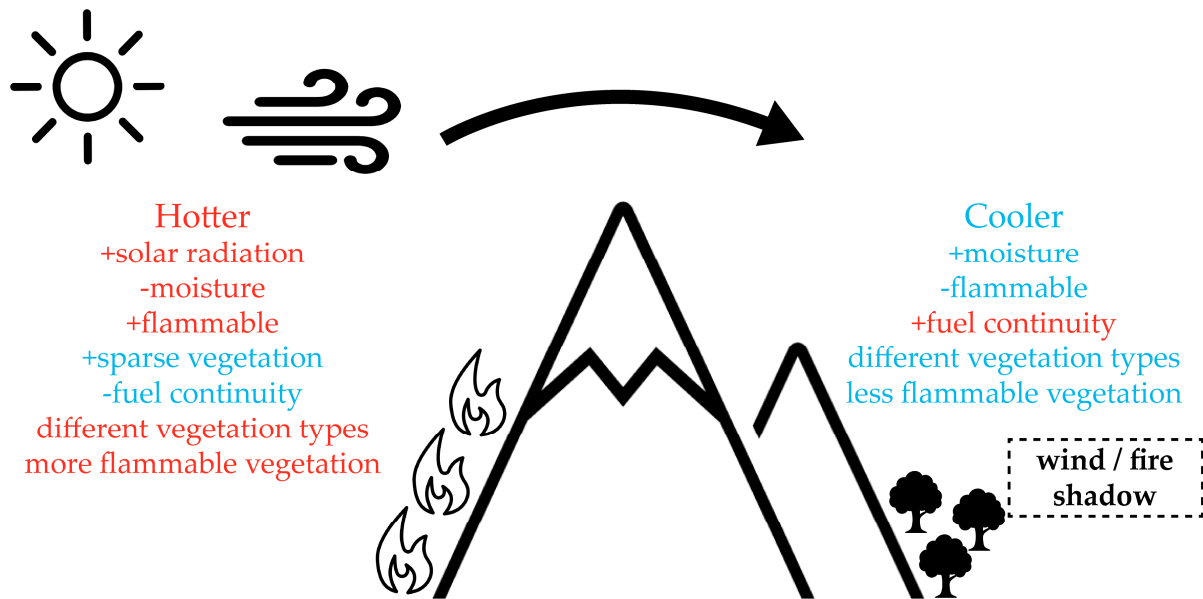


Figure 4. Relationship between aspect, solar radiation (heat budget), moisture balance, vegetation flammability, and predominant fire wind direction, which facilitates wind- and fire-shadows in the lee of the slope. Features that increase fire risk and spread are in red text and those that reduce fire risk and retard fire spread are in blue text. Sun-facing aspect is the north/northwestern slopes in the southern hemisphere.

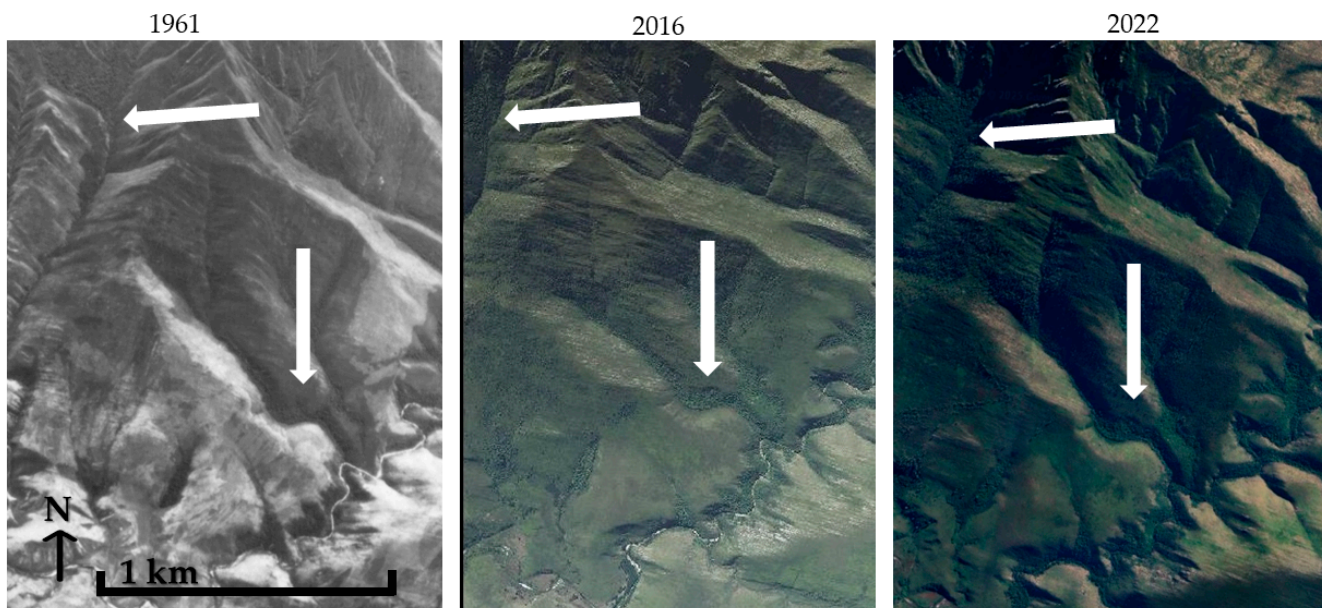


Figure 5. Small, persistent forest fire refugia (indicated by white arrows) in deep valley bottoms on the south-to-southeast slopes. These slopes are in the lee of the predominant fire wind direction from the northeast.

Solar irradiation (particularly global irradiation) is also identified as important across all models and agrees with the findings of [4]. Elevation has previously been identified as important in mapping forest refugia [13], with fire survival increasing at higher elevations. Elevation did not appear to improve model accuracy in this study area. This might be due to the smaller altitudinal range in our study area (0–1675 m) compared to that in the Canadian Rockies (0–3000+ m, [13]) or the southwestern United States (3000–5000 m, [7]) or the Pacific Cordillera of Canada [3]. In addition, the effect of elevation might be emphasised in those studies due to the higher latitudes of the North American forests, which will further

magnify the cooling effect of altitude. Differences in factors identified as important for forest fire refugia will also depend on whether studies focused on persistent forest fire refugia alone, as this study did, or also included ephemeral forest fire refugia (as in the case of [3,7]).

During the last 50 years, climate change seems to have impacted some facets of fire occurrence in the study area. Fire danger weather has become more severe, with higher temperatures and wind speeds and lower relative humidity. This is manifesting in shorter fire return intervals and increasing occurrence of large and high-severity fires [28,29,35,67]. High-severity fires are likely to cause retraction of forest patches, as it has been shown that high-severity fires burn deeper into forest margins [22,24]. However, in the Tsitsikamma area, longer-term rainfall fluctuations over the past ~250 years [27] using stem sections of *Outeniqua* Yellowwood trees collected from the Witelsbos Forest [33] correspond with observations of forest retraction and expansion over longer periods. Forests can expand into fire pathways during wetter, fire-free intervals and then retract to long-term boundaries following severe fires. This pattern is evident across various moist, dry, and scrub forest types in Africa [33]. Potential changes in wind directions due to climate change have not been assessed but are unlikely to affect the fire-flow patterns around and across natural topographic features of the landscape.

4.2. Methods for Mapping Persistent Forest Fire Refugia

Ensemble learning methods and XGBoost consistently achieve the highest accuracy of fit. The stated aims of ensembles are reducing overfitting and enhancing generalisability. KNN did not benefit from ADASYN oversampling, which may highlight its ability to handle unbalanced training data well, as indicated by [61]. However, KNN's performance metrics are the lowest of all models. Across all experiments, choice of variables included seemed to have more impact on model accuracy than the choice of algorithm. We found that it is useful to compare a range of suitably designed models (methods and data), because after aspect, the order of importance of variables was sensitive to model methods (algorithm, use of ADASYN oversampling) and data.

4.3. Limitations

While this study demonstrates the growing capacity of machine learning models to map persistent forest fire refugia, several methodological and data-related constraints should be acknowledged to contextualise the findings and guide future exploration.

As is often the case with machine learning models, data availability and preprocessing represented some of the most significant constraints in this study. Mixed spatial resolution was a particular challenge. To ensure model compatibility, all raster datasets were standardised to the same resolution and extent. Lower-resolution datasets, such as solar irradiation, were upsampled to match the higher-resolution DEM-derived datasets to avoid losing detail from more finely resolved datasets. This approach is commonly applied in geospatial modelling to ensure spatial consistency and preserve key terrain features [47,50]. However, it can introduce artificial precision and minor spatial inconsistencies, particularly in heterogeneous landscapes. The main implication is a potential smoothing of local variability, which may affect the delineation of smaller refugia.

Like many environmental and medical datasets, class imbalance posed a challenge, with substantially more non-refugia than refugia pixels. Such imbalance can bias classification performance toward the majority class. To reduce this ADASYN oversampling, a widely used technique for addressing class imbalance was applied [54]. While ADASYN improved model training by increasing representation of the minority class, more balanced or stratified samples would further strengthen the robustness of future analyses. Additional

techniques, such as ensemble balancing or cost-sensitive learning, may also help reduce residual imbalance effects.

Some model accuracy values may reflect, at least in part, the influence of spatial autocorrelation between training and validation samples. Although PCA preprocessing helped reduce dimensionality and may have mitigated some spatial dependence. We used block holdout to compensate for spatial dependency. Future work could look at minimising spatial clustering in the training data.

5. Conclusions

We found that persistent forest fire refugia can be accurately predicted over large areas and in topographically complex settings, such as in the Sedgefield–Knysna region of southern Africa. Machine learning algorithms, especially ensemble models, using topographic and microclimatic variables emphasise the importance of aspect and its relations to site-level heat and water budgets, in agreement with the findings of [3,13]. Aspect is also related to the occurrence of fire shields from the prevailing fire weather wind direction, as was highlighted, in agreement with [9] in Australia. The importance of surface wind direction has been emphasised by other studies, such as the ‘topographic concavities’ of [7]. While aspect is likely to be important in many areas, the specific aspect will depend on an individual region’s predominant fire wind direction as well as solar exposure, the latter being determined by the hemisphere that a study area is in.

These findings support previous studies emphasising the influences of topography, microclimate, and surface wind on forest fire refugia persistence, particularly in the Sedgefield–Knysna region [5,19]. The machine learning approach used here to predict fire and wind refugia may be applied to other regions where wind and topography are determinants of fire occurrence and vegetation persistence.

In the Sedgefield–Knysna region, these results can be used to highlight the need to protect indigenous forest patches occurring in the lee of topographic features from the prevailing wind direction. These locations are likely to be able to sustain forest in the long term. While climate change may see an increase in fire intensity and frequency, these are unlikely to affect the fire-flow patterns around and across natural topographic features of the landscape. Management of these forest areas may be prioritised, such as to protect these from clear-felling, conversion to agriculture or life-style blocks, or replanting with commercial plantations. These areas can furthermore be incorporated as natural fire breaks or low-risk areas in fire management plans. On the other hand, the identification of areas in the landscape that are not fire refugia may be equally useful to guide fire and vegetation management practices in the landscape. In such fire-exposed areas, the conservation of fire-prone vegetation should be prioritised. These areas should thus be included in prescribed burning schedules, while invasion of forest species [32] and forest establishment efforts should be discouraged here. Use of fire-exposed areas for plantations, as is currently often carried out, is likely to lead to high economic loss.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijgi14120480/s1>, Appendix S1. Patterns based validation, Appendix S2. Fire model pattern based validation, Appendix S3. PCA, Appendix S4. Random forest script with grid search and ADASYN sampling, Appendix S5. XGBoost implementation in Python, Appendix S6. The KNN implementation was performed in Python, Appendix S7. Ensemble script soft, Appendix S8. Ensemble hard script, Appendix S9. Experimental design showing experiment number, dataset and oversampling, X indicate dataset was used in experiments, Appendix S10. Best parameters per experiment.

Author Contributions: Conceptualisation, Sven Christ, Helen M. de Klerk and Coert J. Geldenhuys; methodology, Sven Christ and Helen M. de Klerk; formal analysis, Sven Christ; writing—original draft preparation, Sven Christ and Helen M. de Klerk; writing—review and editing, Helen M. de Klerk, Coert J. Geldenhuys and Tineke Kraaij. All authors have read and agreed to the published version of the manuscript.

Funding: Sven Christ would like to acknowledge the National Institute for the Humanities and Social Sciences scholarship (NIHSS SDS17/1130, <http://nihss.ac.za/>) for funding. The content of this article is solely the responsibility of the authors and does not necessarily represent the views of the funder.

Data Availability Statement: Codes used are supplied in the Appendices in Supplementary Materials. Open-source data are used, and links to public archives thereof are supplied.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Meddens, A.J.H.; Kolden, C.A.; Lutz, J.A.; Smith, A.M.S.; Cansler, C.A.; Abatzoglou, J.T.; Meigs, G.W.; Downing, W.M.; Krawchuk, M.A. Fire Refugia: What Are They, and Why Do They Matter for Global Change? *BioScience* **2018**, *68*, 944–954. [[CrossRef](#)]
2. Román-Cuesta, R.M.; Gracia, M.; Retana, J. Factors influencing the formation of unburned forest islands within the perimeter of a large forest fire. *For. Ecol. Manag.* **2009**, *258*, 71–80. [[CrossRef](#)]
3. Krawchuk, M.A.; Haire, S.L.; Coop, J.; Parisien, M.A.; Whitman, E.; Chong, G.; Miller, C. Topographic and fire weather controls of fire refugia in forested ecosystems of northwestern North America. *Ecosphere* **2016**, *7*, e01632. [[CrossRef](#)]
4. Rogeau, M.-P.; Barber, Q.E.; Parisien, M.-A. Effect of Topography on Persistent Fire Refugia of the Canadian Rocky Mountains. *Forests* **2018**, *9*, 285. [[CrossRef](#)]
5. Geldenhuys, C.J. Bergwind Fires and the Location Pattern of Forest Patches in the Southern Cape Landscape, South-Africa. *J. Biogeogr.* **1994**, *21*, 49–62. [[CrossRef](#)]
6. Giddey, B.L.; Baard, J.A.; Kraaij, T. Verification of the differenced Normalised Burn Ratio (dNBR) as an index of fire severity in Afrotropical Forest. *S. Afr. J. Bot.* **2022**, *146*, 348–353. [[CrossRef](#)]
7. Rodman, K.C.; Davis, K.T.; Parks, S.A.; Chapman, T.B.; Coop, J.D.; Iniguez, J.M.; Roccaforte, J.P.; Meador, A.J.S.; Springer, J.D.; Stevens-Rumann, C.S.; et al. Refuge-yeah or refuge-nah? Predicting locations of forest resistance and recruitment in a fiery world. *Glob. Change Biol.* **2023**, *29*, 7029–7050. [[CrossRef](#)]
8. Rothermel, R.C. *A Mathematical Model for Predicting Fire Spread in Wildland Fuels*; USDA Forest Services Research Paper; Research Paper INT-115; USDA: Washington, DC, USA, 1972; pp. 1–41.
9. Penman, T.D.; Smith, A.; Burton, J.; Heap, A.; McColl-Gausden, S.C.; Najera-Umana, J.; Gordon, F.; Holyland, B.; Marshall, E. What does it take to survive? An expert elicitation approach to understanding the drivers of fire Refugia occurrence and persistence. *Biol. Conserv.* **2025**, *308*, 111257. [[CrossRef](#)]
10. Camp, A.; Oliver, C.; Hessburg, P.; Everett, R. Predicting late-successional fire refugia pre-dating European settlement in the Wenatchee Mountains. *For. Ecol. Manag.* **1997**, *95*, 63–77. [[CrossRef](#)]
11. Chafer, C.J.; Noonan, M.; Macnaught, E. The post-fire measurement of fire severity and intensity in the Christmas 2001 Sydney wildfires. *Int. J. Wildland Fire* **2004**, *13*, 227–240. [[CrossRef](#)]
12. Riley, S.J.; DeGloria, S.D.; Elliot, R. A Terrain Ruggedness Index that Quantifies Topographic Heterogeneity. *Intermt. J. Sci.* **1999**, *5*, 23–27.
13. Rogeau, M.P.; Armstrong, G.W. Quantifying the effect of elevation and aspect on fire return intervals in the Canadian Rocky Mountains. *For. Ecol. Manag.* **2017**, *384*, 248–261. [[CrossRef](#)]
14. Azedou, A.; Lahssini, S.; Khattabi, A.; Meliho, M.; Rifai, N. A Methodological Comparison of Three Models for Gully Erosion Susceptibility Mapping in the Rural Municipality of El Faïd (Morocco). *Sustainability* **2021**, *13*, 682. [[CrossRef](#)]
15. Manders, P.T. Fire and Other Variables as Determinants of Forest Fynbos Boundaries in the Cape-Province. *J. Veg. Sci.* **1990**, *1*, 483–490. [[CrossRef](#)]
16. Coetsee, C.; Bond, W.J.; Wigley, B.J. Forest and fynbos are alternative states on the same nutrient poor geological substrate. *S. Afr. J. Bot.* **2015**, *101*, 57–65. [[CrossRef](#)]
17. Cramer, M.D.; Power, S.C.; Belev, A.; Gillson, L.; Bond, W.J.; Hoffman, M.T.; Hedin, L.O. Are forest-shrubland mosaics of the Cape Floristic Region an example of alternate stable states? *Ecography* **2019**, *42*, 717–729. [[CrossRef](#)]
18. Lu, M.Z.; Bond, W.J.; Sheffer, E.; Cramer, M.D.; West, A.G.; Allsopp, N.; February, E.C.; Chimphango, S.; Ma, Z.Q.; Slingsby, J.A.; et al. Biome boundary maintained by intense belowground resource competition in world's thinnest-rooted plant community. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2117514119. [[CrossRef](#)]

19. Mucina, L.; Geldenhuys, C. Afrotropical, subtropical and azonal forests. In *The Vegetation of South Africa, Lesotho and Swaziland*; Rutherford, L.M.M., Ed.; South African National Biodiversity Institute: Pretoria, South Africa, 2006; pp. 586–615.
20. Pausas, J.G. Alternative fire-driven vegetation states. *J. Veg. Sci.* **2015**, *26*, 4–6. [[CrossRef](#)]
21. Pausas, J.G.; Bond, W.J. Humboldt and the reinvention of nature. *J. Ecol.* **2019**, *107*, 1031–1037. [[CrossRef](#)]
22. Giddey, B.; Baard, J.A.; Vhengani, L.; Kraaij, T. The effect of adjacent vegetation on fire severity in Afrotropical forest along the southern Cape coast of South Africa. *South For.* **2021**, *83*, 225–230. [[CrossRef](#)]
23. Bond, W.J.; Midgley, G.F.; Woodward, F.I. What controls South African vegetation—Climate or fire? *S. Afr. J. Bot.* **2003**, *69*, 79–91. [[CrossRef](#)]
24. Van Wilgen, B.W.; Higgins, K.B.; Bellstedt, D.U. The role of vegetation structure and fuel chemistry in excluding fire from forest patches in the fire-prone fynbos shrublands of South Africa. *J. Ecol.* **1990**, *78*, 210–222. [[CrossRef](#)]
25. Watson LH, C.M. The influence of fire on a southern cape mountain forest. *S. Afr. For. J.* **2001**, *191*, 39–42. [[CrossRef](#)]
26. Strydom, T.; Kraaij, T.; Grobler, B.A.; Cowling, R.M. Canopy plant composition and structure of Cape subtropical dune thicket are predicted by the levels of fire exposure. *PeerJ* **2022**, *10*, e14310. [[CrossRef](#)] [[PubMed](#)]
27. McNaughton, J.; Tyson, P.O. A preliminary assessment of *Podocarpus falcatus* in dendrochronological and dendroclimatological studies in the Witelsbos Forest Reserve. *S. Afr. For. J.* **1979**, *111*, 29–33. [[CrossRef](#)]
28. Kraaij, T.; Baard, J.; Schutte-Vlok, A.L. Plant response to the fire regime (1970–2023) in a fynbos World Heritage Site: Ecological indicators for fire management. *Ecol. Indic.* **2025**, *170*, 113001. [[CrossRef](#)]
29. Kraaij, T.; Baard, J.A.; Cowling, R.M.; van Wilgen, B.W.; Das, S. Historical fire regimes in a poorly understood, fire-prone ecosystem: Eastern coastal fynbos. *Int. J. Wildland Fire* **2013**, *22*, 277–287. [[CrossRef](#)]
30. Quick, L.J.; Chase, B.M.; Wündsche, M.; Kirsten, K.L.; Chevalier, M.; Mäusbacher, R.; Meadows, M.E.; Haberzettl, T. A high-resolution record of Holocene climate and vegetation dynamics from the southern Cape coast of South Africa: Pollen and microcharcoal evidence from Eilandvlei. *J. Quat. Sci.* **2018**, *33*, 487–500. [[CrossRef](#)]
31. Goldblatt, P.; Manning, J. *Cape Plants: A Conspectus of the Cape Flora of South Africa*; National Botanical Institute: Pretoria, South Africa; Missouri Botanical Garden: St Louis, MO, USA, 2000; Volume 9.
32. Geldenhuys, C.J.; Moll, E.J.; Swart, R.C. Evergreen forests in South Africa—Their composition, biogeography, ecological dynamics and sustainable use management. *S. Afr. J. Bot.* **2026**, *188*, 63–98. [[CrossRef](#)]
33. Geldenhuys, C.J. *Composition and Dynamics of Plant Communities in the Southern Cape Forests*; CSIR, Division of Forest Science and Technology: Pretoria, South Africa, 1993; p. 56.
34. Geldenhuys, C.J. The Management of the Southern Cape Forests. *S. Afr. For. J.* **1982**, *121*, 4–10. [[CrossRef](#)]
35. Kraaij, T.; Cowling, R.M.; van Wilgen, B.W. Lightning and fire weather in eastern coastal fynbos shrublands: Seasonality and long-term trends. *Int. J. Wildland Fire* **2013**, *22*, 288–295. [[CrossRef](#)]
36. Giglio, L.; Boschetti, L.; Roy, D.P.; Humber, M.L.; Justice, C.O. The Collection 6 MODIS burned area mapping algorithm and product. *Remote Sens. Environ.* **2018**, *217*, 72–85. [[CrossRef](#)]
37. Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; et al. The shuttle radar topography mission. *Rev. Geophys.* **2007**, *45*, RG2004. [[CrossRef](#)]
38. NASA JPL. NASA Shuttle Radar Topography Mission Global 1 arc Second [Data Set]. 2013. Available online: <https://www.earthdata.nasa.gov/data/catalog/lpcloud-srtmg1-003> (accessed on 4 February 2023). [[CrossRef](#)]
39. Elkhachy, I. Vertical accuracy assessment for SRTM and ASTER Digital Elevation Models: A case study of Najran city, Saudi Arabia. *Ain Shams Eng. J.* **2018**, *9*, 1807–1817. [[CrossRef](#)]
40. Mashimbye, Z.E.; de Clercq, W.P.; Van Niekerk, A. An evaluation of digital elevation models (DEMs) for delineating land components. *Geoderma* **2014**, *213*, 312–319. [[CrossRef](#)]
41. Huggel, C.; Schneider, D.; Miranda, P.J.; Delgado Granados, H.; Käab, A. Evaluation of ASTER and SRTM DEM data for lahar modeling: A case study on lahars from Popocatepetl Volcano, Mexico. *J. Volcanol. Geotherm. Res.* **2008**, *170*, 99–110. [[CrossRef](#)]
42. Grass Development Team. *Geographic Resources Analysis Support System (GRASS) Software*, Version 8.4; Open Source Geospatial Foundation: Arlington, VA, USA, 2024.
43. Solargis. *Solar Resource Map*; Solargis: Singapore, 2021; Available online: <https://solargis.com/> (accessed on 28 November 2025).
44. Biljecki, F.; Heuvelink, G.B.M.; Ledoux, H.; Stoter, J. Propagation of positional error in 3D GIS: Estimation of the solar irradiation of building roofs. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 2269–2294. [[CrossRef](#)]
45. Christ, S. *The Influence of Topographical Variability on Wildfire Occurrence and Propagation*; Stellenbosch University: Stellenbosch, South Africa, 2024.
46. Morvan, D. Physical Phenomena and Length Scales Governing the Behaviour of Wildfires: A Case for Physical Modelling. *Fire Technol.* **2011**, *47*, 437–460. [[CrossRef](#)]
47. Sharples, J.J.; McRae, R.H.; Wilkes, S.R. Wind–terrain effects on the propagation of wildfires in rugged terrain: Fire channelling. *Int. J. Wildland Fire* **2012**, *21*, 282–296. [[CrossRef](#)]

48. Frost, P.; Kleyn, L.; Van Den Dool, R.; Burgess, M.; Vhengani, L.; Steenkamp, K.; Wessels, K. *The Elandskraal Fire, Knysna: A Data Driven Analysis*; CSIR Meraka Institute: Pretoria, South Africa, 2018.
49. Gallagher, C.A.; Chudzinska, M.; Larsen-Gray, A.; Pollock, C.J.; Sells, S.N.; White, P.J.C.; Berger, U. From theory to practice in pattern-oriented modelling: Identifying and using empirical patterns in predictive models. *Biol. Rev.* **2021**, *96*, 1868–1888. [[CrossRef](#)] [[PubMed](#)]
50. Pithani, M.B.; Sanyal, S.; Shukla, A.K. Bilinear and Bicubic Interpolations for Image Presentation of Mechanical Stress and Temperature Distribution. *Power Eng. Eng. Thermophys.* **2022**, *1*, 8–18. [[CrossRef](#)]
51. Gao, Z.T.; Wang, L.M.; Wu, G.S. LIP: Local Importance-based Pooling. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV 2019), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3354–3363.
52. Chandra, R.; Bansal, C.; Kang, M.Y.; Blau, T.; Agarwal, V.; Singh, P.; Wilson, L.O.W.; Vasan, S. Unsupervised machine learning framework for discriminating major variants of concern during COVID-19. *PLoS ONE* **2023**, *18*, e0285719. [[CrossRef](#)]
53. He, H.; Bai, Y.; Garcia, E.A.; Li, S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *IEEE Int. Jt. Conf. Neural Netw.* **2008**, *2008*, 1322–1328. [[CrossRef](#)]
54. Ahmed, G.; Er, M.J.; Fareed, M.M.S.; Zikria, S.; Mahmood, S.; He, J.; Asad, M.; Jilani, S.F.; Aslam, M. DAD-Net: Classification of Alzheimer’s Disease Using ADASYN Oversampling Technique and Optimized Neural Network. *Molecules* **2022**, *27*, 7085. [[CrossRef](#)]
55. Cutler, D.R.; Edwards, T.; Beard, K.H.; Cutler, A.; Hess, K.T. Random forests for classification in ecology. *Ecology* **2007**, *88*, 2783–2792. [[CrossRef](#)]
56. Ramos, D.; Carneiro, D.; Novais, P. evoRF: An Evolutionary Approach to Random Forests. *Stud. Comput. Intell.* **2020**, *868*, 102–107. [[CrossRef](#)]
57. Chen, X.; Ishwaran, H. Random forests for genomic data analysis. *Genomics* **2012**, *99*, 323–329. [[CrossRef](#)] [[PubMed](#)]
58. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *KDD 16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; Krishnapuram, B., Shah, M., Eds.; Association for Computing Machinery: New York, NY, USA; San Francisco, CA, USA, 2016; pp. 785–794.
59. Georganos, S.; Grippa, T.; Vanhuyse, S.; Lennert, M.; Shimoni, M.; Wolff, E. Very High Resolution Object-Based Land Use-Land Cover Urban Classification Using Extreme Gradient Boosting. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 607–611. [[CrossRef](#)]
60. Ma, C.L.; Peng, Y.X.; Wu, L.T.; Guo, X.Y.; Wang, X.B.; Kong, X.Q. Application of Machine Learning Techniques to Predict the Occurrence of Distraction-affected Crashes with Phone-Use Data. *Transp. Res. Rec.* **2022**, *2676*, 692–705. [[CrossRef](#)]
61. Thanh Noi, P.; Kappas, M. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors* **2017**, *18*, 18. [[CrossRef](#)]
62. Prakash, K.B.; Kanagachidambaresan, G.R. Pattern Recognition and Machine Learning. In *Programming with TensorFlow; EAI/Springer Innovations in Communication and Computing*; Prakash, K.B., Kanagachidambaresan, G.R., Eds.; Springer: Cham, Switzerland, 2021; pp. 105–144.
63. Barzani, A.R.; Pahlavani, P.; Ghorbanzadeh, O.; Gholamnia, K.; Ghamisi, P. Evaluating the Impact of Recursive Feature Elimination on Machine Learning Models for Predicting Forest Fire-Prone Zones. *Fire Ecol.* **2024**, *7*, 440. [[CrossRef](#)]
64. Zhang, L.; Shi, C.; Zhang, F. Predicting Forest Fire Area Growth Rate Using an Ensemble Algorithm. *Forests* **2024**, *15*, 1493. [[CrossRef](#)]
65. Tantithamthavorn, C.; McIntosh, S.; Hassan, A.E.; Matsumoto, K. The Impact of Automated Parameter Optimization on Defect Prediction Models. *IEEE Trans. Softw. Eng.* **2019**, *45*, 683–711. [[CrossRef](#)]
66. Goldstein, B.A.; Polley, E.C.; Briggs, F.B.S. Random Forests for Genetic Association Studies. *Stat. Appl. Genet. Mol.* **2011**, *10*, 32. [[CrossRef](#)] [[PubMed](#)]
67. Kraaij, T.; Baard, J.A.; Arndt, J.; Vhengani, L.; van Wilgen, B.W. An assessment of climate, weather, and fuel factors influencing a large, destructive wildfire in the Knysna region, South Africa. *Fire Ecol.* **2018**, *14*, 4. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.